STA 213: INTRODUCTION TO STATCRUNCH

StatCrunch is a computer-driven statistical package that will be used to perform most statistical analyses discussed in class. StatCrunch may be used on **any** computer that has access to the Internet. Using StatCrunch requires two essential elements: (1) Entering data into a spread-sheet, and (2) Requesting some analysis be done with the data values entered, usually the construction of a graph or some form of computation.

*Caution: Please remember that there will be a few things you are expected to be able to do without the aid of a computer. Class notes would highlight these. Additionally, please remember that **<u>you are always expected to explain and interpret any computer output that you obtain</u>**. Use complete sentences and place results in the context of the problem!*

ACCESSING STATCRUNCH

StatCrunch is available on *any* computer that is connected to the Internet. The URL for the program is:

http://access.nku.edu/statcrunch/

The program requires that your browser is able to use JAVA. Additionally, if you see a security question, click "**Yes**" (or "**Always**") to accept the security certificate. If you have problems accessing StatCrunch, I *may* be able to provide some measure of assistance if you can explain your problem clearly in an email.

The program also requires a **passcode** for you to use it. The passcode is "**heyu**" (note that it is case-sensative). Type these four characters in the passcode box at the bottom of the screen, and click "Submit". After a few moments, a worksheet window similar to Excel should appear.

ENTERING DATA INTO STATCRUNCH

Before you can utilize StatCrunch's capabilities, you must have data entered into the worksheet. There are two standard ways that will be used to enter data into the StatCrunch worksheet.

For most textbook problems, you can **load data from the internet**. The method of doing this is as follows:

1. Choose **Data → Load Data → From WWW** (note: these are menu selections from the menus at the top)

2. In the WWW address box, type **http://www.nku.edu/~statistics/data213/Ch##/*filename*.xls.** Note that ## must be replaced by the appropriate chapter number and *filename* is replaced by the appropriate data file name given in the textbook or homework problem. (Note: If looking at review materials from 212, those files are accessed using: **http://www.nku.edu/~statistics/data212/*filename*.xls.** In rare cases I will provide you with a data set that is not from the text, and the location for those would have "~statistics" replaced by "~nolanj1". Please note that the web address is case-sensitive and must be entirely lower case.

3. Click "OK" and the data should now appear in the spreadsheet.

Loading data **from a local file** is similar. Depending on the type of file, you may need to change the "delimiter" in order for the file to load correctly. It is also possible (but also somewhat painful) to enter data into the worksheet **by typing it yourself**. How to do this is fairly self explanatory. Simply "click" the cursor into the appropriate box and type. Note that unfortunately the only way to remove characters is "backspace" (or "delete" to clear a cell). You may change variable names as well as enter data. You may also adjust variable names for a downloaded data set.
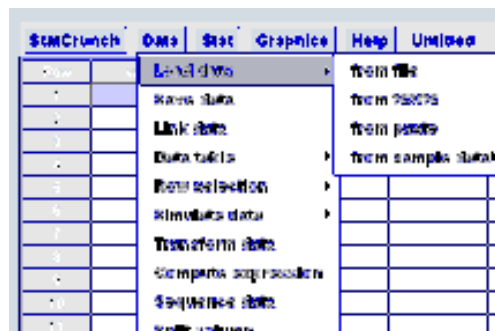
USING STATCRUNCH FOR GRAPHICS AND COMPUTATIONS

Throughout the class, StatCrunch will be used to create graphs and perform computations. In fact, only a very few of these operations would we ask you to perform by hand. Your goals should be to

- learn to use StatCrunch to get what you need

- learn to copy items from SC and paste them into MSWord or another appropriate editor. This includes resizing them appropriately so that you can then discuss the output (see homework solutions for ideas as to what this should look like)

- learn to take the results from StatCrunch and interpret them in the context of a problem

- learn to communicate your conclusions and the reasons for them (verbally and/or in writing)

Details on how to use StatCrunch will be given in class, with handouts similar to this posted to Blackboard.

COPYING AND/OR EXPORTING RESULTS

If an analysis requires you to use StatCrunch, you should show the StatCrunch output. Output from StatCrunch can easily be pasted into a Word document. With the appropriate analysis window open, click **Options → Copy**. The contents of that window have been copied onto the clipboard and can now be pasted into Word. Oftentimes, I find that using the paste-special command in word, and pasting the output as a picture (enhanced metafile) is quite useful even for text output. This allows the output to be easily resized and moved around within the document.

Often during the semester I will ask you to do a small analysis, write it up, and turn it in as an activity or homework. You should do this in Word, and submit your document including **both output and explanation**.

HISTOGRAMS IN STATCRUNCH

In Chapter 2, we discuss visual representations of data. In particular, we focus on the construction and interpretation of a histogram. Fortunately, StatCrunch will construct histograms for us (hence you are not asked to construct them by hand).

Before you can construct a histogram, you need to load data into StatCrunch using one of the methods above. Remember that the data you will use for a histogram needs to be **quantitative**. As an example, we will load the dataset STUDENTS.xls (from STA 212). If you have properly loaded this data, it should look like this:

Notice that there are both quantitative and qualitative variables in this data set. The procedure for constructing a histogram for any quantitative variable is as follows:

1. From the menu, select **Graphics → Histogram.** This yields the box shown in (a) below.

2. Click on the appropriate **quantitative** variable for which you want to construct a histogram (note that non-numerical variables will not even appear). Your variable of choice will appear in the gray box on the right. Then click **Next**.

    o Note: You may wish to group by another variable by selecting a *qualitative* variable (e.g. gender) from the **"Group by"** menu. For example, grouping by "gender" would produce one histogram for males and another for females. If you do this, you will want to check the box for "separate graph for each group".

3. In the next box, shown in (b), first select the appropriate type of histogram to make – either frequency (count) or relative frequency (percentage). You may optionally select a starting value and bin width (or perhaps more appropriately you may come back to these after seeing your first attempt). Click **Next** twice, skipping the screen that allows you to choose a "density".

4. You will now see the box shown in (c). You should *always* make sure your graphs/axes are labeled appropriately. If you are creating multiple histograms at once (for example, using groups), the next button will allow you to put them into the same picture (leave the "title" blank if you do this, so the group variable will appear in the title). Otherwise, click **Create Graph** and your histogram will appear.

|		
(a) Step 2.	**(b) Step 3.**	**(c) Step 4.**

Note that there are additional options that I have not given here – I encourage you to "play" with them to see what they do. Once you construct a histogram, it will appear in a new window. You can choose **Options → Edit** to change some of your options. You can also choose **Options → Copy** to place the histogram on your clipboard so that it may be pasted into Word. Always remember that once you have a histogram, you should interpret the histogram using complete sentences and addressing those criteria we covered in class.

A sample histogram for age, grouped by gender, is shown below:
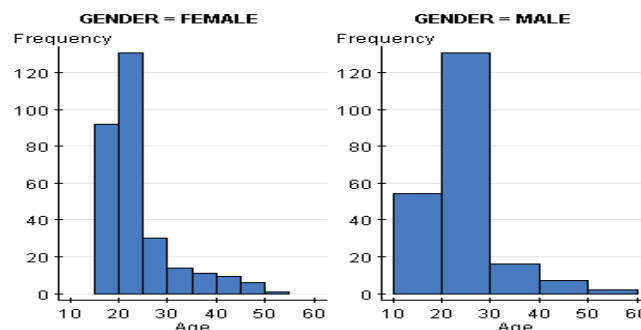
OTHER GRAPHICAL ANALYSES

At times we may also want to consider other graphical analyses. The mechanics of using these in StatCrunch are fairly similar to that of a histogram – so I will not discuss each in detail. I will, however, provide a list of the graphical functions that you may wish to use at some point during the class:

- Bar Plot (or bar graph): Appropriate to visually display either categorical (qualitative) or ordinal data.
- Pie Chart: Another option for displaying categorical or ordinal data.
- Box Plot: A good option for displaying quantitative data, especially when comparing groups (use the group by categorical variable functionality).
- Scatter Plot: Used to display the relationship between two quantitative variables (see regression)
- QQ Plot: Used to assess the normality of a variable (or, in regression, of the residuals). Finding a nearly straight line formed in this plot suggests that an assumption of normality is probably ok.
- Index Plot: Plots the data against the order of operations. Useful if the order in which observations were collected may cause some pattern in the data (time series).

COMPUTING SUMMARY STATISTICS IN STATCRUNCH

In Chapter 3, we discuss numerical summary statistics such as the mean, median, and standard deviation. These are also relatively easy to obtain using StatCrunch. With data loaded as before:

1. From the menu, select **Stat → Summary Stats → Columns.** This yields the box shown in (a) below.

2. Click quantitative variables to select them (they will appear in the grey box at the right when selected). Note that you may use "grouping" in the same manner as for histograms. Click **Next**.

3. Using the box shown in (b), you may select / deselect the statistics you wish to compute. The options you will use most often include:

   - n = number of observations
   - Mean = sample mean
   - Std. Dev. = sample standard deviation
   - Std. Err. = standard error of the sample mean
   - Median = 50th percentile

4. Click **Calculate** and the statistics for all variables you selected will pop up in a table.

As before, you can edit the options from the results window, and copy the results for pasting into Word. Sample results are shown below:

Column	n	Mean	Std. Dev.	Min	Median	Max
AGE	504	23.311508	6.6482325	17	21	56
WORKHR	503	24.595427	14.022032	0	25	60
BARHR	504	1.4623016	2.9861455	0	0	25

If you have questions about SC, you are welcome to come by my office! The next tutorial in this series illustrates how to use SC to do the one and two-sample tests.

Last Updated January 2010